《Generative Adversarial Nets》

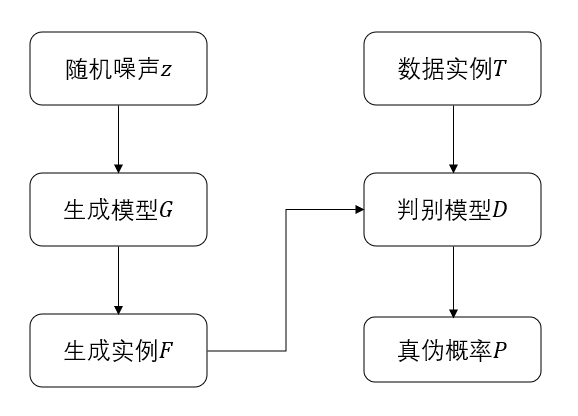
**摘要：**

1. **论文背景：**深度学习已经在判别模型中取得了一系列较好的成果，但在生成模型领域中尚未有优秀、合适的框架被提出。
2. **论文的贡献：**提出了一个通过对抗过程来估计生成模型的新神经网络框架，它包含两个训练模型：**模拟数据分布的生成模型G**和**判别样本数据真伪的判别模型D**。
3. **主要创新点：**所提出的架构可以有效使用反向传播来进行训练，而不需要运用任何类似“马尔科夫链”复杂的推理网络。此外，框架还结合了一定的博弈理论。
4. **实验分析：**论文中简单证明了对抗性建模框架（**生成对抗网络**）的可行性，并且表明了这些研究方向可能是有用的。【该框架如今已被广泛使用】

**GAN网络整体框架：**

在GAN中包含有两个模型，一个是**生成模型G（Generative model）**，一个是**判别模型D（Discriminative model）**。生成模型G的任务是生成看起来自然真实的、和原始数据相似的实例；而判别模型D的任务是判断给定的实例看起来是自然真实的，还是人为伪造的（真实实例来源于数据集，伪造实例来源于生成模型G）。

具体网络结构如图所示：



1. **生成模型G：**G是一个生成伪造实例的网络，它首先接收一个随机的噪声，然后通过这个噪声来生成伪造（生成）实例，记做。
2. **判别模型D：**D是一个用于判别的网络，其主要功能就是判别一个实例是不是“真实的”。它的输入是（代表一个实例），输出是，它代表为真实实例的概率。如果，则代表100% 是真实实例；而如果，则代表不可能是真实实例。

**注意：**这里的网络是指MLP（多层感知器）。

**GAN模型优化训练：**

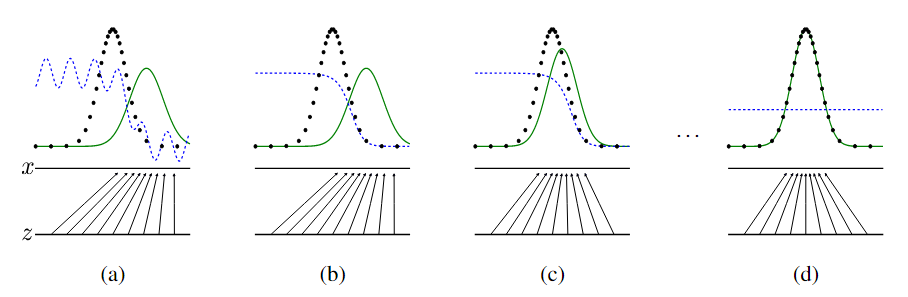
GAN的目标优化函数如下：



训练的目标就是优化判别网络，使得最大概率地分对训练样本的标签（即最大化和)）；优化生成网络，使得最小化，即最大化的损失。

而在训练过程中，首先固定一方网络，然后再更新另一个网络的参数，并交替迭代，使得对方的错误最大化【类似于博弈过程】。最终，生成模型G就能估测出样本数据的分布，即使得生成的样本更加真实。

**对抗过程可视化：**



如图所示，在模型的对抗训练中，可能会出现以上几种情况：

**图(a)：**D（蓝色线）刚开始训练，其本身判别能力有限，结果存在波动，但还是能初步区分实际数据和生成数据。

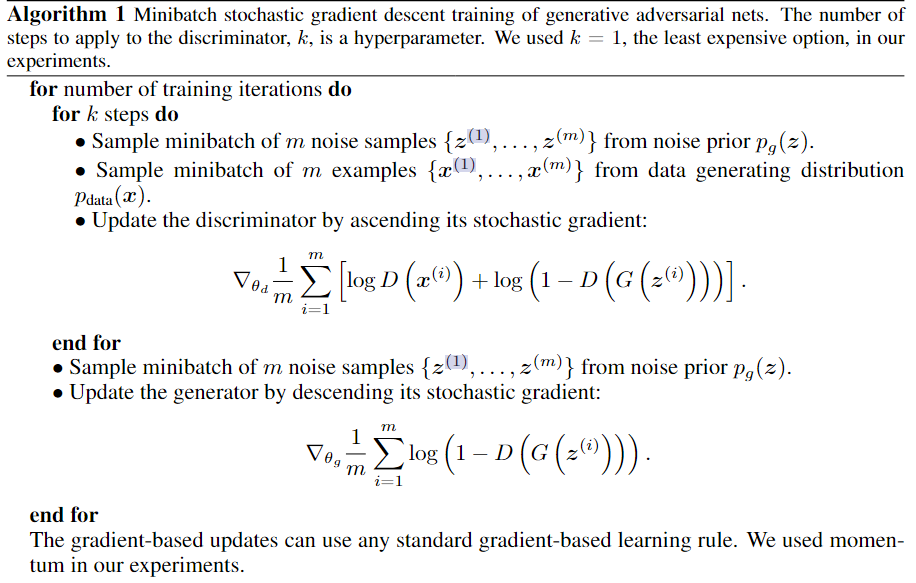
**图(b)：**在D经过有效训练后，可以明显地区分出生成数据。

**图(c)：**G优化自己的输出分布（绿色线），使得其更加接近真实数据的分布（黑色线），从而使得D更加难以判别实际数据和生成数据。

**图(d)：**由于G的不断提升，其输出分布几乎已经等同于真实数据的分布，从而使得D难以判别二者，因此D的输出结果趋近于。至此，网络G和D处于**纳什均衡状态**，无法再进一步优化。

**整体算法的实现步骤：**

如图所示：



首先，用步优化判别模型D的参数。然后，固定D，优化生成模型G的参数（采用反向传播来优化网络参数）。以上过程进行次迭代。

**一些数学证明：**